

**PSY 652 Lab: Module 3 Activity**  
**September 18, 2019**

The following exercises are intended to provide hands-on practice conceptualizing how different types of Questionable Research Practices (QRPs) impact study results and biases in published literature. While many types of QRPs exist, here we'll specifically focus on two types of HARKing: Cherry Picking and Question Trolling. We've included two tables from Murphy & Aguinis' 2019 article that used simulations to examine how Cherry Picking and Question Trolling impact statistical results. Answer the questions below using the tables and provided terms.

*Note: Without HARKing, the true parameter for the average correlation coefficient (mean  $r$ ) across studies is .20 in these simulations.*

**Definitions of key terms:**

- 1.) **QRPs:** Questionable Research Practices
- 2.) **HARKing:** Hypothesizing After Results are Known
- 3.) **Forms of HARKing:**
  - a.) Cherry Picking: Involves searching through data that includes alternative measures or samples to find the results that offer the strongest possible support for a particular hypothesis or research question a study was designed to investigate. *Tends to be associated with a more homogenous set of effects in the published literature.*
  - b.) Question Trolling: Involves searching through data that includes several different constructs, interventions, or relationships to find seemingly notable results. *Tends to be associated with more heterogeneous sets of effects in the published literature.*
- 4.) **Sample Size:** The number of participants included in a study.
- 5.) **Pool Size:** The total number of statistical tests a research team conducts and chooses from when selecting results to publish (e.g., if a researcher conducts 10 tests and elects to only publish the one test that detected the largest effect size, pool size = 10; if a researcher conducts 2 tests and elects to only publish the test that detected a larger effect size, pool size = 2).
- 6.) **QRP Prevalence:** The percentage of studies that employed a specific type of QRP. In the studies included below, QRP prevalence was specified as a simulation parameter.
- 7.) **Mean  $r$ :** The average correlation coefficient reported across a set of studies. In this case, mean  $r$  is our proxy for effect size, where larger  $r$  values reflect larger effect sizes.
- 8.) **Question Trolling Heterogeneity (aka effect size SD):** In this context, effect size SD represents the upper bound for variability in effect sizes in each set of studies. The more haphazardly a research team trolls for questions (i.e., the more variable relationships they examine), the more variability they will observe in effect sizes (i.e., a wider range of effect sizes to choose from when selecting results to publish).

**Section 1: Impact of cherry picking results**

*Below are results from Murphy & Aguinis' (2019) HARKing simulation.*

**Table 2** HARKed estimates of the population correlation

		Pool size				
Number		2	4	6	8	10
Cherry-picking						
Prevalence						
20%	100	0.212	0.219	0.223	0.226	0.229
	140	0.209	0.217	0.220	0.223	0.225
	180	0.208	0.215	0.218	0.220	0.222
	220	0.207	0.213	0.216	0.218	0.219
	260	0.207	0.212	0.215	0.217	0.218
	280	0.206	0.212	0.214	0.216	0.217
40%	100	0.223	0.239	0.246	0.252	0.258
	140	0.218	0.233	0.241	0.246	0.249
	180	0.216	0.229	0.235	0.240	0.244
	220	0.215	0.227	0.232	0.236	0.239
	260	0.213	0.224	0.229	0.233	0.237
	280	0.212	0.224	0.229	0.233	0.234
60%	100	0.235	0.258	0.269	0.278	0.287
	140	0.226	0.250	0.261	0.269	0.274
	180	0.224	0.244	0.253	0.260	0.265
	220	0.222	0.240	0.248	0.254	0.258
	260	0.220	0.237	0.244	0.250	0.255
	280	0.218	0.235	0.243	0.249	0.252
80%	100	0.247	0.277	0.293	0.304	0.316
	140	0.235	0.267	0.282	0.292	0.299
	180	0.232	0.258	0.271	0.280	0.287
	220	0.229	0.253	0.264	0.272	0.278
	260	0.226	0.249	0.258	0.267	0.273
	280	0.224	0.247	0.258	0.265	0.269

- 1) Conceptually, what does this table demonstrate?
    - a) Why does mean  $r$  tend to increase with increasing pool size?
    - b) Why does mean  $r$  tend to decrease with increasing sample size?
    - c) What is the impact of increasing cherry picking prevalence?
-

**Section 2: Question Trolling**

*Below are results from Murphy & Aguinis' (2019) HARKing simulation.*

		Number	Pool size				
			2	4	6	8	10
Question trolling							
Heterogeneity							
SD = 0.05							
20%	100	0.213	0.222	0.226	0.230	0.232	
	140	0.210	0.220	0.224	0.227	0.228	
	180	0.210	0.218	0.222	0.225	0.226	
	220	0.210	0.216	0.220	0.223	0.225	
	260	0.209	0.216	0.220	0.222	0.220	
	280	0.209	0.216	0.219	0.222	0.220	
40%	100	0.225	0.243	0.253	0.261	0.264	
	140	0.220	0.240	0.247	0.253	0.257	
	180	0.219	0.235	0.243	0.249	0.252	
	220	0.219	0.233	0.241	0.247	0.250	
	260	0.218	0.232	0.239	0.244	0.246	
	280	0.218	0.231	0.239	0.243	0.247	
60%	100	0.238	0.265	0.279	0.291	0.297	
	140	0.230	0.259	0.271	0.280	0.285	
	180	0.229	0.253	0.265	0.274	0.279	
	220	0.229	0.249	0.261	0.270	0.276	
	260	0.226	0.247	0.259	0.266	0.269	
	280	0.227	0.247	0.258	0.265	0.270	
80%	100	0.251	0.287	0.305	0.321	0.329	
	140	0.240	0.279	0.294	0.307	0.314	
	180	0.238	0.271	0.286	0.299	0.305	
	220	0.238	0.266	0.281	0.293	0.301	
	260	0.235	0.263	0.279	0.287	0.293	
	280	0.236	0.263	0.277	0.286	0.294	

**Table 2** HARKed estimates of the population correlation

		Number	Pool size				
			2	4	6	8	10
SD = 0.10							
20%	100	0.217	0.230	0.237	0.241	0.245	
	140	0.216	0.228	0.235	0.239	0.242	
	180	0.216	0.228	0.234	0.238	0.241	
	220	0.215	0.227	0.233	0.237	0.239	
	260	0.215	0.226	0.232	0.237	0.239	
	280	0.215	0.227	0.233	0.236	0.240	
40%	100	0.234	0.260	0.273	0.283	0.290	
	140	0.231	0.256	0.270	0.277	0.285	
	180	0.232	0.255	0.267	0.277	0.281	
	220	0.229	0.253	0.266	0.274	0.279	
	260	0.230	0.253	0.264	0.274	0.278	
	280	0.231	0.254	0.265	0.273	0.279	
60%	100	0.250	0.290	0.310	0.324	0.334	
	140	0.247	0.285	0.305	0.316	0.327	
	180	0.247	0.283	0.301	0.315	0.322	
	220	0.244	0.280	0.299	0.311	0.318	
	260	0.245	0.279	0.296	0.311	0.318	
	280	0.246	0.281	0.298	0.309	0.319	
80%	100	0.267	0.320	0.347	0.365	0.379	
	140	0.262	0.313	0.340	0.355	0.370	
	180	0.263	0.310	0.334	0.353	0.362	
	220	0.259	0.307	0.332	0.348	0.358	
	260	0.260	0.306	0.328	0.347	0.357	
	280	0.262	0.308	0.331	0.345	0.359	

**Table 2** HARKed estimates of the population correlation

		Pool size				
Number		2	4	6	8	10
SD = 0.15						
20%	100	0.220	0.238	0.246	0.252	0.254
	140	0.219	0.237	0.244	0.250	0.254
	180	0.220	0.237	0.245	0.249	0.254
	220	0.219	0.237	0.243	0.248	0.253
	260	0.220	0.236	0.242	0.249	0.253
	280	0.219	0.235	0.243	0.247	0.252
40%	100	0.239	0.275	0.292	0.303	0.309
	140	0.238	0.274	0.287	0.299	0.309
	180	0.239	0.274	0.289	0.298	0.307
	220	0.238	0.274	0.287	0.297	0.305
	260	0.239	0.272	0.285	0.297	0.306
	280	0.237	0.270	0.285	0.294	0.305
60%	100	0.259	0.313	0.337	0.355	0.363
	140	0.258	0.310	0.331	0.349	0.363
	180	0.259	0.311	0.334	0.347	0.361
	220	0.257	0.311	0.330	0.345	0.358
	260	0.259	0.308	0.327	0.346	0.359
	280	0.256	0.305	0.328	0.342	0.357
80%	100	0.279	0.351	0.383	0.407	0.417
	140	0.277	0.347	0.375	0.399	0.417
	180	0.279	0.345	0.378	0.398	0.413
	220	0.276	0.348	0.374	0.394	0.411
	260	0.278	0.343	0.370	0.395	0.410
	280	0.275	0.340	0.371	0.389	0.409

2. Conceptually, what does this table demonstrate?

- a. Why does mean  $r$  tend to increase with increasing pool size?
- b. Why does mean  $r$  tend to decrease with increasing sample size?
- c. What is the impact of increasing question trolling prevalence?
- d. Why is it useful to include Question Trolling Heterogeneity (effect size SD) in the question trolling simulation?
- e. What happens as the Question Trolling Heterogeneity (effect size SD) increases?

3. What four factors affect the impact of bias produced from Cherry Picking & Question Trolling?

4. Based on these tables, is Cherry Picking or Question Trolling more likely to be introducing biases into published literature? How did you reach your conclusion?

5. What are three ways that researchers, editors, and journals can respond to QRPs and the current “reproducibility crisis” in published literature?